International Journal of Recent Research and Review, Vol. VIII, Issue 2, June 2015

ISSN 2277 – 8322

Performance Analysis of Association Rule Mining Techniques

Gaurav Mathur[1], Chetan Kumar[2], Sandeep K. Gupta[3]

[1, 2] *Department of Computer Science and Engineering , Kautilya Institute of Technology & Engineering, Jaipur, India*

[3]I*CFAI University, Jaipur, India*

Email: gauravm.adr@gmail.com

***Abstract*—Association rule mining can be acknowledged as unsupervised learning model which explores the interesting relationship and impressions among large set of data items on the basis of magnitude of some predefined threshold. Support-confidence based association rule mining is the traditional paradigm and process in the association rule mining which employs confidence for final rule formation from all maximum probable rules. In this paper, analysis as well as study of performance of Apriori technique and Boolean matrix technique has been carried out.**

***Keywords*—Data mining, Association rules mining, Apriori, Review, Analysis.**

I. INTRODUCTION

Association rules mining catches interesting associations or relations between items or item sets in an abundant data container [1]. While mining process large number of rules gets come to existence but only the small set has significance regarding user perspective, so to find and analysis whether a rule is of user interest or not it must require a relevant metric to weigh the degree of rule in which the user is passionate. Thus interestingness measure plays an important role in rule extraction process. This notion was first introduced by Agrawal [2] to analyze the data of a supermarket containing large collection of customer transaction. It is S.C. Satapathy et al. a kind of unsupervised learning technique [3]. The following is the broad scheme of an association rule:

$$X \Rightarrow Y$$

In the above rule, X is recognized as antecedent and Y is recognized as consequent and it can be dictated as if X then Y. It signifies the association between X and Y. Support-confidence model for association rule can be described as given: Let I = i1 , i2 , ..., in be a set of n literals called items and T be a set of transactions, where each transaction t ∈ T is a set of items such that t ⊆ I. Every exclusive transaction is accomplice with a unique id TID. An association rule is the reference of the form, X ⇒ Y, where X ⊆ I, Y ⊆ I and X ∩Y = ∅ [4]. Here X is the antecedent and Y is the consequent. Terms used in rule mining process are characterized as follows:

A. Support (sup)

Support of an itemset can be formalized as the ratio of transactions accommodating the items in both antecedent and consequent of the rule to the total number of transaction [5]. It weighs the stability of the given itemset. Let X ⇒ Y be the rule then support is characterized as:

$$\text{sup}(X \Rightarrow Y) = \text{sup count}(X \cap Y)/T \quad (1)$$

Here, sup count is the number of transactions accommodating the given itemset X∩Y.

B. Confidence (conf)

Confidence of an association rule determines whether and how often items in Y appear in transaction that posses X [6]. It weighs the strength of a given association rule. Let X ⇒ Y be the rule then confidence is characterized as:

$$\text{conf}(X \Rightarrow Y) = \text{sup}(X \cup Y)/\text{sup}(X) \quad (2)$$

The rule mining task gets carried out in two steps:

(i) Frequent Itemset Identification

The itemset that amuse the minimum support threshold (σ) gets discovered or generated in this phase.

(ii) Rule Extraction

This phase takes frequent itemset as input which has been accomplished in phase 1 along with the magnitude of minimum confidence threshold (τ). The rules which delight the minimum confidence are extracted in this phase.

II. Related Work

1) *Apriori ARM Technique*

Apriori algorithm is a prominent algorithm to extract association rules. Apriori implements iterative methods like Layer by Layer Search. Apriori can be segregate as a seminal algorithm. The algorithm has been backed by the fact that this algorithm utilizes prior knowledge domain of frequent item sets. This algorithm employs an iterative level wise search approach so that k – item sets can be used to explore (k+1) – item sets. First, frequent 1 -item sets is obtained which is denoted as L1. The frequency of each item in L1 delights the minimum support. After wards then L1 is use to calculate L2 which is the set of frequent 2 – item sets, L2 is use to render L3 until no frequent k – item sets can be explored [6].

2) *Boolean ARM Technique*

Boolean Matrix is a yet fast known algorithm that employs Boolean relational calculus to explore unknown impression between items of a transactional database. The key factor that bought edge to this algorithm is no requirement of precious candidate formation, only single time database examination and accumulation of data in bit form to lessen memory cost. These association rules can be formatted as one-dimensional or multidimensional lying upon the number of predicates [7]. For example buys (A, “Camera”) ⇒ buys (A, “Printer”) which express 'A' the one who buys Camera also buys Printer, is a single dimensional as it only posses’ one predicate “buys”. But how so ever, age(A, "30") ∧ occupation (A, "manager") ⇒ buys (A, "mobile"), which express one who is 30 years of age and who is manager buys a mobile, is multidimensional as it contains three predicates “age”, “occupation” and “buys” . Multidimensional rule are also characterized as inter-dimensional rule [8].

III. Experiments & results

The experiments of above algorithms are conducted on machine with Intel(R) Core I-5 processor having 2.60 GHz processor speed and 4 GB occupied RAM memory. The operating system used is Windows-8. The relational transactional database used has 10000 records and 20000 records respectively for separate experiments with single dimension dataset. Database rows denote transactions and columns denote transaction items or products. The data is stored in 0’s and 1’s form that is bit form where 0 represent absence of item in the cell for that particular row transaction and 1 represent its presence. The experimental results of Apriori and Boolean matrix are shown in table I and performance analysis is shown in figure1.

TABLE I

COMPARISONS OF TIME TAKEN BY APRIORI AND BOOLEAN MATRIX OVER TEN AND TWENTY THOUSAND RECORDS

S. No	Number of Records	Apriori Time Taken (ms)	Boolean Matrix Time Taken (ms)
1	10000	1354	143
2	20000	2026	168

As presented in the table, for ten thousand records Apriori takes 1354 ms while Boolean Matrix requires only 143 ms, same goes for twenty thousand records where Apriori takes 2026 ms in contrast to 168 ms taken by Boolean Matrix.

In the graph, X axis shows number of records while Y axis shows the time in milliseconds taken by Apriori and Boolean Matrix algorithms respectively.

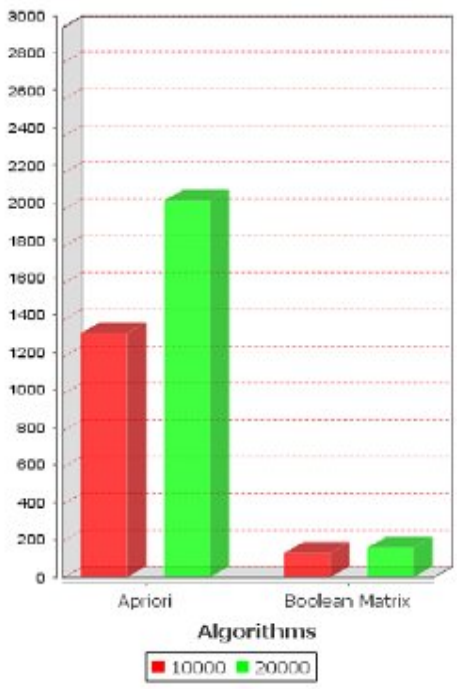

Fig. 1 Graph of time taken in ms by Apriori and Boolean Matrix for ten and twenty thousand records respectively.

IV. Analysis

To extract and likewise mine association rules whether multidimensional or single dimensional from a relational database, Apriori algorithm is utilized which translates the database into transaction database and then iteratively examines the database to find correlation among the items that is very precious in relevance to memory and time. To wipe out such obstacles Boolean Matrix has been utilized. Boolean matrix does examine the database only ones so the time consume by Boolean matrix technique is far less than Apriori technique.

V. Conclusion

The leading peculiar features of the Boolean matrix are that it examines the database only once, it does not un-necessary creates the candidate item sets, and it uses the Boolean vector to render frequent item sets. It accumulate all data in bits, so it demand very tiny memory space and can be employed to large relational databases while Apriori examines the dataset iteratively and hence require abundant time in contrast to Boolean matrix. The results also show that Boolean matrix is faster than Apriori technique.

VI. References

[1] LUO XianWen, WANG WeiQing, "*Improved Algorithms Research for Association Rule Based on Matrix*", 2010 International Conference on Intelligent Computing and Cognitive Informatics, IEEE, 2010.

[2] Agrawal R, Lmielinski T, Swami A. "*Mining Association Rules between Sets of Items in Large Database*", Proceeding of the ACM SIGMOD Conference on Management of Data, Washington, USA, pp. 207-216, 1993.

[3] Agrawal, R., Srikant, R.: "*Fast algorithm for mining association rules.*" In: Proceeding of 20th International Conference on Very Large Databases, pp. 487–499 (2003).

[4] Pratima Gautam, K. R. Pardasani, "*A Fast Algorithm for Mining Multilevel Association Rule Based on Boolean Matrix*", International Journal on Computer Science and Engineering, Vol. 02, No. 03, page no. 746-752, 2010.

[5] Xuezhi Chi, "*A New Matrix-Based Association Rules Mining Algorithm"*, 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012).

[6] Imielinski, T., Agrawal, R., Swami, A.N.: "*Mining association rules between sets of items in large databases*". In: Proceedings of the ACM SIGMOD Conference on Management of Data, vol. 22, pp. 207–216 (1993).

[7] Reda ALHAJJ, Mehmet KAYA, "*Integrating Fuzziness into OLAP for Multidimensional Fuzzy Association Rules Mining*",Canada Third IEEE International Conference on Data Mining (ICDM'0 (ICDM'03).

[8] Neelu Khare1, Neeru Adlakha2, K. R. Pardasani3: "*An Algorithm for Mining Multidimensional Association Rules using Boolean Matrix*", 2010 International Conference on Recent Trends in Information, Telecommunication and Computing.